



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 15/12, C07K 14/47, C12N 15/10, 15/11</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 99/06550</b> <b>(43) International Publication Date:</b> 11 February 1999 (11.02.99)
<b>(21) International Application Number:</b> PCT/IB98/01232 <b>(22) International Filing Date:</b> 31 July 1998 (31.07.98) <b>(30) Priority Data:</b> 08/905,144      1 August 1997 (01.08.97)      US <b>(71) Applicant (for all designated States except US):</b> GENSET [FR/FR]; 24, rue Royale, F-75008 Paris (FR). <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> DUMAS MILNE EDWARDS, Jean-Baptiste [FR/FR]; 8, rue Grégoire-de-Tours, F-75006 Paris (FR). DUCLERT, Aymeric [FR/FR]; 6 ter, rue Victorine, F-94100 Saint-Maur (FR). LACROIX, Bruno [FR/FR]; 93, route de Vourles, F-69230 Saint-Genis Laval (FR). <b>(74) Agents:</b> MARTIN Jean-Jacques et al.; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).		<b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
<b>(54) Title:</b> 5' ESTs FOR SECRETED PROTEINS EXPRESSED IN PROSTATE  <b>(57) Abstract</b>  <p>The sequences of 5' ESTs derived from mRNAs encoding secreted proteins are disclosed. The 5' ESTs may be to obtain cDNAs and genomic DNAs corresponding to the 5' ESTs. The 5' ESTs may also be used in diagnostic, forensic, gene therapy, and chromosome mapping procedures. Upstream regulatory sequences may also be obtained using the 5' ESTs. The 5' ESTs may also be used to design expression vectors and secretion vectors.</p>		

have been developed to isolate human promoters. One of them consists of making a CpG island library (Cross, *et al.*, *Nature Genetics* 6: 236-244, 1994). The second consists of isolating human genomic DNA sequences containing SpeI binding sites by the use of SpeI binding protein. (Mortlock *et al.*, *Genome Res.* 6:327-335, 1996). Both of these approaches  
5 have their limits due to a lack of specificity or of comprehensiveness.

The present 5' ESTs may be used to efficiently identify and isolate upstream regulatory regions which control the location, developmental stage, rate, and quantity of protein synthesis, as well as the stability of the mRNA. (Theil, *BioFactors* 4:87-93, 1993). Once identified and characterized, these regulatory regions may be utilized in gene therapy or  
10 protein purification schemes to obtain the desired amount and locations of protein synthesis or to inhibit, reduce, or prevent the synthesis of undesirable gene products.

In addition, ESTs containing the 5' ends of secretory protein genes may include sequences useful as probes for chromosome mapping and the identification of individuals. Thus, there is a need to identify and characterize the sequences upstream of the 5' coding  
15 sequences of genes encoding secretory proteins.

#### Summary of the Invention

The present invention relates to purified, isolated, or recombinant ESTs which include sequences derived from the authentic 5' ends of their corresponding mRNAs. The term  
20 "corresponding mRNA" refers to the mRNA which was the template for the cDNA synthesis which produced the 5' EST. These sequences will be referred to hereinafter as "5' ESTs." As used herein, the term "purified" does not require absolute purity; rather, it is intended as a relative definition. Individual 5' EST clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from these  
25 clones could not be obtained directly either from the library or from total human DNA. The cDNA clones are not naturally occurring as such, but rather are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The conversion of mRNA into a cDNA library involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection. Thus,  
30 creating a cDNA library from messenger RNA and subsequently isolating individual clones from that library results in an approximately  $10^4$ - $10^6$  fold purification of the native message.

Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

As used herein, the term "isolated" requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide present in a living animal is not isolated, but the same polynucleotide, separated from some or all of the coexisting materials in the natural system, is isolated.

As used herein, the term "recombinant" means that the 5' EST is adjacent to "backbone" nucleic acid to which it is not adjacent in its natural environment. Additionally, to be "enriched" the 5' ESTs will represent 5% or more of the number of nucleic acid inserts in a population of nucleic acid backbone molecules. Backbone molecules according to the present invention include nucleic acids such as expression vectors, self-replicating nucleic acids, viruses, integrating nucleic acids, and other vectors or nucleic acids used to maintain or manipulate a nucleic acid insert of interest. Preferably, the enriched 5' ESTs represent 15% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. More preferably, the enriched 5' ESTs represent 50% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. In a highly preferred embodiment, the enriched 5' ESTs represent 90% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules.

"Stringent", moderate, and "low" hybridization conditions are as defined in Example 29.

Unless otherwise indicated, a "complementary" sequence is fully complementary.

Thus, 5' ESTs in cDNA libraries in which one or more 5' ESTs make up 5% or more of the number of nucleic acid inserts in the backbone molecules are "enriched recombinant 5' ESTs" as defined herein. Likewise, 5' ESTs in a population of plasmids in which one or more 5' EST of the present invention have been inserted such that they represent 5% or more of the number of inserts in the plasmid backbone are "enriched recombinant 5' ESTs" as defined herein. However, 5' ESTs in cDNA libraries in which 5' ESTs constitute less than 5% of the number of nucleic acid inserts in the population of backbone molecules, such as libraries in

which backbone molecules having a 5' EST insert are extremely rare, are not "enriched recombinant 5' ESTs."

In particular, the present invention relates to 5' ESTs which are derived from genes encoding secreted proteins. As used herein, a "secreted" protein is one which, when expressed in a suitable host cell, is transported across or through a membrane, including transport as a result of signal peptides in its amino acid sequence. "Secreted" proteins include without limitation proteins secreted wholly (e.g. soluble proteins), or partially (e.g. receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins which are transported across the membrane of the endoplasmic reticulum.

Such 5' ESTs include nucleic acid sequences, called signal sequences, which encode signal peptides which direct the extracellular secretion of the proteins encoded by the genes from which the 5' ESTs are derived. Generally, the signal peptides are located at the amino termini of secreted proteins.

Secreted proteins are translated by ribosomes associated with the "rough" endoplasmic reticulum. Generally, secreted proteins are co-translationally transferred to the membrane of the endoplasmic reticulum. Association of the ribosome with the endoplasmic reticulum during translation of secreted proteins is mediated by the signal peptide. The signal peptide is typically cleaved following its co-translational entry into the endoplasmic reticulum.

After delivery to the endoplasmic reticulum, secreted proteins may proceed through the Golgi apparatus. In the Golgi apparatus, the proteins may undergo post-translational modification before entering secretory vesicles which transport them across the cell membrane.

The 5' ESTs of the present invention have several important applications. For example, they may be used to obtain and express cDNA clones which include the full protein coding sequences of the corresponding gene products, including the authentic translation start sites derived from the 5' ends of the coding sequences of the mRNAs from which the 5' ESTs are derived. These cDNAs will be referred to hereinafter as "full length cDNAs." These cDNAs may also include DNA derived from mRNA sequences upstream of the translation start site. The full length cDNA sequences may be used to express the proteins corresponding to the 5' ESTs. As discussed above, secreted proteins are therapeutically important. Thus, the proteins expressed from the cDNAs may be useful in treating or

controlling a variety of human conditions. The 5' ESTs may also be used to obtain the corresponding genomic DNA. The term "corresponding genomic DNA" refers to the genomic DNA which encodes the mRNA from which the 5' EST was derived.

Alternatively, the 5' ESTs may be used to obtain and express extended cDNAs  
5 encoding portions of the secreted protein. The portions may comprise the signal peptides of the secreted proteins or the mature proteins generated when the signal peptide is cleaved off. The portions may also comprise polypeptides having at least 10 consecutive amino acids encoded by the extended cDNAs or full length cDNAs. Alternatively, the portions may comprise at least 15 consecutive amino acids encoded by the extended cDNAs or full length  
10 cDNAs. In some embodiments, the portions may comprise at least 25 consecutive amino acids encoded by the extended cDNAs or full length cDNAs. In other embodiments, the portions may comprise at least 40 amino acids encoded by the extended cDNAs or full length cDNAs.

Antibodies which specifically recognize the entire secreted proteins encoded by the  
15 extended cDNAs, full length cDNAs, or fragments thereof having at least 10 consecutive amino acids, at least 15 consecutive amino acids, at least 25 consecutive amino acids, or at least 40 consecutive amino acids may also be obtained as described below. Antibodies which specifically recognize the mature protein generated when the signal peptide is cleaved may also be obtained as described below. Similarly, antibodies which specifically recognize the  
20 signal peptides encoded by the extended cDNAs or full length cDNAs may also be obtained.

In some embodiments, the extended cDNAs obtained using the 5' ESTs include the signal sequence. In other embodiments, the extended cDNAs obtained using the 5' ESTs may include the full coding sequence for the mature protein (*i.e.* the protein generated when the signal polypeptide is cleaved off). In addition, the extended cDNAs obtained using the 5'  
25 ESTs may include regulatory regions upstream of the translation start site or downstream of the stop codon which control the amount, location, or developmental stage of gene expression.

As discussed above, secreted proteins are therapeutically important. Thus, the proteins expressed from the extended cDNAs or full length cDNAs obtained using the 5'  
30 ESTs may be useful in treating or controlling a variety of human conditions.

The 5' ESTs (or cDNAs or genomic DNAs obtained therefrom) may be used in forensic procedures to identify individuals or in diagnostic procedures to identify individuals having genetic diseases resulting from abnormal expression of the genes corresponding to the 5' ESTs. In addition, the present invention is useful for constructing a high resolution map of the human chromosomes.

The present invention also relates to secretion vectors capable of directing the secretion of a protein of interest. Such vectors may be used in gene therapy strategies in which it is desired to produce a gene product in one cell which is to be delivered to another location in the body. Secretion vectors may also facilitate the purification of desired proteins.

The present invention also relates to expression vectors capable of directing the expression of an inserted gene in a desired spatial or temporal manner or at a desired level. Such vectors may include sequences upstream of the 5' ESTs, such as promoters or upstream regulatory sequences.

Finally, the present invention may also be used for gene therapy to control or treat genetic diseases. Signal peptides may also be fused to heterologous proteins to direct their extracellular secretion.

Bacterial clones containing Bluescript plasmids having inserts containing the 5' ESTs of the present invention (SEQ ID NOs: 38-315 are presently stored at 80°C in 4% (v/v) glycerol in the inventor's laboratories under the designations listed next to the SEQ ID NOs in II). The inserts may be recovered from the deposited materials by growing the appropriate clones on a suitable medium. The Bluescript DNA can then be isolated using plasmid isolation procedures familiar to those skilled in the art such as alkaline lysis minipreps or large scale alkaline lysis plasmid isolation procedures. If desired the plasmid DNA may be further enriched by centrifugation on a cesium chloride gradient, size exclusion chromatography, or anion exchange chromatography. The plasmid DNA obtained using these procedures may then be manipulated using standard cloning techniques familiar to those skilled in the art. Alternatively, a PCR can be done with primers designed at both ends of the EST insertion. The PCR product which corresponds to the 5' EST can then be manipulated using standard cloning techniques familiar to those skilled in the art.

One aspect of the present invention is a purified or isolated nucleic acid having the sequence of one of SEQ ID NOs: 38-315 or having a sequence complementary thereto. In one embodiment, the nucleic acid is recombinant.

Another aspect of the present invention is a purified or isolated nucleic acid  
5 comprising at least 10 consecutive bases of the sequence of one of SEQ ID NOs: 38-315 or one of the sequences complementary thereto.

Yet another aspect of the present invention is a purified or isolated nucleic acid comprising at least 15 consecutive bases of one of the sequences of SEQ ID NOs: 38-315 or one of the sequences complementary thereto. In one embodiment, the nucleic acid is  
10 recombinant.

A further aspect of the present invention is a purified or isolated nucleic acid of at least 15 bases capable of hybridizing under stringent conditions to the sequence of one of SEQ ID NOs: 38-315 or one of the sequences complementary to the sequences of SEQ ID NOs: 38-315. In one embodiment, the nucleic acid is recombinant.

Another aspect of the present invention is a purified or isolated nucleic acid encoding  
15 a human gene product, said human gene product having a sequence partially encoded by one of the sequences of SEQ ID NO: 38-315.

Still another aspect of the present invention is a method of making a cDNA encoding a human secretory protein, said human secretory protein being partially encoded by one of  
20 SEQ ID NOs 38-315, comprising the steps of contacting a collection of mRNA molecules from human cells with a primer comprising at least 15 consecutive nucleotides of a sequence complementary to one of SEQ ID NOs: 38-315; hybridizing said primer to an mRNA in said collection that encodes said protein; reverse transcribing said hybridized primer to make a first cDNA strand from said mRNA; making a second cDNA strand complementary to said first  
25 cDNA strand; and isolating the resulting cDNA encoding said protein comprising said first cDNA strand and said second cDNA strand.

Another aspect of the invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-315 or a fragment thereof of at least 10 amino acids, said cDNA being  
30 obtainable by the method described in the preceding paragraph. In one embodiment, the

cDNA comprises the full protein coding sequence of said protein which sequence is partially included in one of the sequences of SEQ ID NOs: 38-315.

Another aspect of the present invention is a method of making a cDNA encoding a human secretory protein that is partially encoded by one of SEQ ID NOs 38-315, comprising  
5 the steps of obtaining a cDNA comprising one of the sequences of SEQ ID NOs: 38-315; contacting said cDNA with a detectable probe comprising at least 15 consecutive nucleotides of said sequence of SEQ ID NO: 38-315 or a sequence complementary thereto under conditions which permit said probe to hybridize to said cDNA; identifying a cDNA which hybridizes to said detectable probe; and isolating said cDNA which hybridizes to said probe.

10 Another aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-315 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method described in the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the  
15 sequences of SEQ ID NOs: 38-315.

Another aspect of the present invention is a method of making a cDNA comprising one of the sequence of SEQ ID NOs: 38-315, comprising the steps of contacting a collection of mRNA molecules from human cells with a first primer capable of hybridizing to the polyA tail of said mRNA; hybridizing said first primer to said polyA tail; reverse transcribing said  
20 mRNA to make a first cDNA strand; making a second cDNA strand complementary to said first cDNA strand using at least one primer comprising at least 15 nucleotides of one of the sequences of SEQ ID NOs 38-315; and isolating the resulting cDNA comprising said first cDNA strand and said second cDNA strand.

Another aspect of the present invention is an isolated or purified cDNA encoding a  
25 human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-315 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method described in the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-315.

30 In one embodiment of the method described in the two paragraphs above, the second cDNA strand is made by contacting said first cDNA strand with a first pair of primers, said



first pair of primers comprising a second primer comprising at least 15 consecutive nucleotides of one of the sequences of SEQ ID NOs 38-315 and a third primer having a sequence therein which is included within the sequence of said first primer; performing a first polymerase chain reaction with said first pair of nested primers to generate a first PCR product; contacting said first PCR product with a second pair of primers, said second pair of primers comprising a fourth primer, said fourth primer comprising at least 15 consecutive nucleotides of said sequence of one of SEQ ID NOs: 38-315, and a fifth primer, said fourth and fifth primers being capable of hybridizing to sequences within said first PCR product; and performing a second polymerase chain reaction, thereby generating a second PCR product.

One aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-315, or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-315.

Another aspect of the present invention is the method described four paragraphs above in which the second cDNA strand is made by contacting said first cDNA strand with a second primer comprising at least 15 consecutive nucleotides of the sequences of SEQ ID NOs: 38-315; hybridizing said second primer to said first strand cDNA; and extending said hybridized second primer to generate said second cDNA strand.

Another aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein partially encoded by one of SEQ ID NOs 38-315 or comprising a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method described in the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-315.

Another aspect of the present invention is a method of making a protein comprising one of the sequences of SEQ ID NOs: 316-593, comprising the steps of obtaining a cDNA encoding the full protein sequence partially included in one of the sequences of sequence of SEQ ID NOs: 38-315; inserting said cDNA in an expression vector such that said cDNA is

operably linked to a promoter, introducing said expression vector into a host cell whereby said host cell produces the protein encoded by said cDNA; and isolating said protein.

Another aspect of the present invention is an isolated protein obtainable by the method described in the preceding paragraph.

5 Another aspect of the present invention is a method of obtaining a promoter DNA comprising the steps of obtaining DNAs located upstream of the nucleic acids of SEQ ID NOs: 38-315 or the sequences complementary thereto; screening said upstream DNAs to identify a promoter capable of directing transcription initiation; and isolating said DNA comprising said identified promoter. In one embodiment, the obtaining step comprises  
10 chromosome walking from said nucleic acids of SEQ ID NOs: 38-315 or sequences complementary thereto. In another embodiment, the screening step comprises inserting said upstream sequences into a promoter reporter vector. In another embodiment, the screening step comprises identifying motifs in said upstream DNAs which are transcription factor binding sites or transcription start sites.

15 Another aspect of the present invention is an isolated promoter obtainable by the method described above.

Another aspect of the present invention is an isolated or purified protein comprising one of the sequences of SEQ ID NOs: 316-593.

Another aspect of the present invention is the inclusion of at least one of the  
20 sequences of SEQ ID NOs: 38-315, or one of the sequences complementary to the sequences of SEQ ID NOs: 38-315, or a fragment thereof of at least 15 consecutive nucleotides in an array of discrete ESTs or fragments thereof of at least 15 nucleotides in length. In one embodiment, the array includes at least two of the sequences of SEQ ID NOs: 38-315, the sequences complementary to the sequences of SEQ ID NOs: 38-315, or fragments thereof of  
25 at least 15 consecutive nucleotides. In another embodiment, the array includes at least five of the sequences of SEQ ID NOs: 38-315, the sequences complementary to the sequences of SEQ ID NOs: 38-315, or fragments thereof of at least 15 consecutive nucleotides.

Another aspect of the present invention is a promoter having a sequence selected from the group consisting of SEQ ID NOs: 31, 34, and 37.

CLAIMS

1. A purified or isolated nucleic acid comprising the sequence of one of SEQ ID NOs: 38-315 or comprising a sequence complementary thereto.
- 5 2. The nucleic acid of Claim 1, wherein said nucleic acid is recombinant.
3. A purified or isolated nucleic acid comprising at least 10 consecutive bases of the sequence of one of SEQ ID NOs: 38-315 or one of the sequences complementary thereto.
4. A purified or isolated nucleic acid comprising at least 15 consecutive bases of  
10 one of the sequences of SEQ ID NOs: 38-315 or one of the sequences complementary thereto.
5. The nucleic acid of Claim 4, wherein said nucleic acid is recombinant.
6. A purified or isolated nucleic acid of at least 15 bases capable of hybridizing under stringent conditions to the sequence of one of SEQ ID NOs: 38-315 or one of the  
15 sequences complementary to the sequences of SEQ ID NOs: 38-315.
7. The nucleic acid of Claim 6, wherein said nucleic acid is recombinant.
8. A purified or isolated nucleic acid encoding a human gene product, said human gene product having a sequence partially encoded by one of the sequences of SEQ ID NO: 38-315.
- 20 9. A purified or isolated nucleic acid having the sequence of one of SEQ ID NOs: 38-315 or having a sequence complementary thereto.
10. A purified or isolated nucleic acid comprising the nucleotides of one of SEQ ID NOs: 38-315 which encode a signal peptide.
11. A purified or isolated polypeptides comprising a signal peptide encoded by  
25 one of the sequences of SEQ ID NOs: 38-315.
12. A vector encoding a fusion protein comprising a polypeptide and a signal peptide, said vector comprising a first nucleic acid encoding a signal peptide encoded by one of the sequences of SEQ ID NOs: 38-315 operably linked to a second nucleic acid encoding a polypeptide.
- 30 13. A method of directing the extracellular secretion of a polypeptide or the insertion of a polypeptide into the membrane comprising the steps of:

obtaining a vector according to Claim 12; and

introducing said vector into a host cell such that said fusion protein is secreted into the extracellular environment of said host cell or inserted into the membrane of said host cell.

14. A method of importing a polypeptide into a cell comprising contacting said  
5 cell with a fusion protein comprising a signal peptide encoded by one of the sequences of SEQ ID NOs: 38-315 operably linked to said polypeptide.

15. A method of making a cDNA encoding a human secretory protein that is partially encoded by one of SEQ ID NOs 38-315, comprising the steps of:

obtaining a cDNA comprising one of the sequences of SEQ ID NOs: 38-315;

10 contacting said cDNA with a detectable probe comprising at least 15 consecutive nucleotides of said sequence of SEQ ID NO: 38-315 or a sequence complementary thereto under conditions which permit said probe to hybridize to said cDNA;

identifying a cDNA which hybridizes to said detectable probe; and

isolating said cDNA which hybridizes to said probe.

16. An isolated or purified cDNA encoding a human secretory protein, said  
15 human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-315 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 15.

17. The cDNA of Claim 16 wherein said cDNA comprises the full protein coding  
20 sequence partially included in one of the sequences of SEQ ID NOs: 38-315.

18. A method of making a cDNA comprising one of the sequences of SEQ ID NOs: 38-315, comprising the steps of:

contacting a collection of mRNA molecules from human cells with a first primer capable of hybridizing to the polyA tail of said mRNA;

25 hybridizing said first primer to said polyA tail;

reverse transcribing said mRNA to make a first cDNA strand;

making a second cDNA strand complementary to said first cDNA strand using at least one primer comprising at least 15 nucleotides of one of the sequences of SEQ ID NOs 38-315; and

30 isolating the resulting cDNA comprising said first cDNA strand and said second cDNA strand.

19. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-315 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 18.

5        20. The cDNA of Claim 19 wherein said cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-315.

21. The method of Claim 18, wherein the second cDNA strand is made by:

contacting said first cDNA strand with a first pair of primers, said first pair of primers comprising a second primer comprising at least 15 consecutive nucleotides of one of the sequences of SEQ ID NOs 38-315 and a third primer having a sequence therein which is  
10 included within the sequence of said first primer;

performing a first polymerase chain reaction with said first pair of nested primers to generate a first PCR product;

contacting said first PCR product with a second pair of primers, said second pair of  
15 primers comprising a fourth primer, said fourth primer comprising at least 15 consecutive nucleotides of said sequence of one of SEQ ID NOs 38-315, and a fifth primer, said fourth and fifth primers being capable of hybridizing to sequences within said first PCR product; and  
performing a second polymerase chain reaction, thereby generating a second PCR product.

20        22. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-315, or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 21.

23. The cDNA of Claim 22 wherein said cDNA comprises the full protein coding  
25 sequence partially included in one of the sequences of SEQ ID NOs: 38-315.

24. The method of Claim 18 wherein the second cDNA strand is made by:

contacting said first cDNA strand with a second primer comprising at least 15 consecutive nucleotides of the sequences of SEQ ID NOs: 38-315;

hybridizing said second primer to said first strand cDNA; and

30        extending said hybridized second primer to generate said second cDNA strand.

25. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein partially encoded by one of SEQ ID NOs 38-315 or comprising a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 24.

5        26. The cDNA of Claim 25, wherein said cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-315.

27. A method of making a protein comprising one of the sequences of SEQ ID NO: 316-593, comprising the steps of:

obtaining a cDNA encoding the full protein sequence partially included in one of the  
10 sequences of sequence of SEQ ID NO: 38-315;

inserting said cDNA in an expression vector such that said cDNA is operably linked to a promoter;

introducing said expression vector into a host cell whereby said host cell produces the protein encoded by said cDNA; and

15 isolating said protein.

28. An isolated protein obtainable by the method of Claim 27.

29. A method of obtaining a promoter DNA comprising the steps of:

obtaining DNAs located upstream of the nucleic acids of SEQ ID NO: 38-315 or the sequences complementary thereto;

20 screening said upstream DNAs to identify a promoter capable of directing transcription initiation; and

isolating said DNA comprising said identified promoter.

30. The method of Claim 29, wherein said obtaining step comprises chromosome walking from said nucleic acids of SEQ ID NO: 38-315 or sequences complementary thereto.

25 31. The method of Claim 30, wherein said screening step comprises inserting said upstream sequences into a promoter reporter vector.

32. The method of Claim 30, wherein said screening step comprises identifying motifs in said upstream DNAs which are transcription factor binding sites or transcription start sites.

30 33. An isolated promoter obtainable by the method of Claim 32.

34. An isolated or purified protein comprising one of the sequences of SEQ ID NO: 316-593.

35. In an array of discrete ESTs or fragments thereof of at least 15 nucleotides in length, the improvement comprising inclusion in said array of at least one of the sequences of  
5 SEQ ID NOs: 38-315, or one of the sequences complementary to the sequences of SEQ ID NOs: 38-315, or a fragment thereof of at least 15 consecutive nucleotides.

36. The array of Claim 35 including therein at least two of the sequences of SEQ ID NOs: 38-315, the sequences complementary to the sequences of SEQ ID NOs: 38-315, or fragments thereof of at least 15 consecutive nucleotides.

10 37. The array of Claim 35 including therein at least five of the sequences of SEQ ID NOs: 38-315, the sequences complementary to the sequences of SEQ ID NOs: 38-315, or fragments thereof of at least 15 consecutive nucleotides.

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 75..263  
 (C) IDENTIFICATION METHOD: fasta  
 (D) OTHER INFORMATION: identity 99  
                           region 1..189  
                           id HSCRISP3G  
                           vrt

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 51..146  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 7.7  
                           seq LLFLVAGLLPSFP/AN

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 311:

```

AATATATACG GCTCTAACCT TCTCTCTCTG CACCTTCCTT CTGTCAATAG ATG AAA      56
                                     Met Lys

CAA ATA CTT CAT CCT GCT CTG GAA ACC ACT GCA ATG ACA TTA TTC CCA      104
Gln Ile Leu His Pro Ala Leu Glu Thr Thr Ala Met Thr Leu Phe Pro
-30                -25                -20                -15

GTG CTG TTG TTC CTG GTT GCT GGG CTG CTT CCA TCT TTT CCA GCA AAT      152
Val Leu Leu Phe Leu Val Ala Gly Leu Leu Pro Ser Phe Pro Ala Asn
                -10                -5                1

GAA GAT AAG GAT CCC GCT TTT ACT GCT TTG TTA ACC ACC CAA ACA CAA      200
Glu Asp Lys Asp Pro Ala Phe Thr Ala Leu Leu Thr Thr Gln Thr Gln
                5                10                15

GTG CAA AGG GAG ATT GTG AAT AAG CAC AAT GAA CTG AGG AGA GCA GTA      248
Val Gln Arg Glu Ile Val Asn Lys His Asn Glu Leu Arg Arg Ala Val
                20                25                30

TCT CCC CCT GCC AAA      263
Ser Pro Pro Ala Lys
35

```

## (2) INFORMATION FOR SEQ ID NO: 312:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 465 base pairs  
 (B) TYPE: NUCLEIC ACID  
 (C) STRANDEDNESS: DOUBLE  
 (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
 (F) TISSUE TYPE: Cancerous prostate

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 133..467